

Speedup of lexicographic optimization by superiorization and its applications to cancer radiotherapy treatment

Esther Bonacker¹, Aviv Gibali², Karl-Heinz Küfer¹
and Philipp Süss¹

September 12, 2016

¹ Optimization Department, Fraunhofer - ITWM, Kaiserslautern 67663, Germany

² Mathematics Department, ORT Braude College, Karmiel 2161002, Israel

E-mail: esther.bonacker@itwm.fraunhofer.de and avivg@braude.ac.il

Abstract

Multicriteria optimization problems occur in many real life applications, for example in cancer radiotherapy treatment and in particular in intensity modulated radiation therapy (IMRT). In this work we focus on optimization problems with multiple objectives that are ranked according to their importance. We solve these problems numerically by combining lexicographic optimization with our recently proposed level set scheme, which yields a sequence of auxiliary convex feasibility problems; solved here via projection methods. The projection enables us to combine the newly introduced superiorization methodology with multicriteria optimization methods to speed up computation while guaranteeing convergence of the optimization. We demonstrate our scheme with a simple 2D academic example (used in the literature) and also present results from calculations on four real head neck cases in IMRT (Radiation Oncology of the Ludwig-Maximilians University, Munich, Germany) for two different choices of superiorization parameter sets suited to yield fast convergence for each case individually or robust behavior for all four cases.

Keywords: Superiorization, projection methods, feasibility problems, multi-criteria optimization, iterative methods

MSC: 65K10, 65K15, 90C25, 90C90

1 Introduction

A wide variety of real-life problems involve the simultaneous optimization of several objectives or criteria. Objectives that contradict each are common, thus resulting in situations where one objective cannot be improved without worsening at least one other objective. For example, in cancer radiotherapy treatment, one goal is to maximize the amount of cancerous tissue exposed to radiation treatment while another (contradicting) goal is to minimize the radiation exposure of healthy organs. Here, there is typically no trivial compromise. Problems of this kind are called *Multi-Criteria Optimization* (MCO) problems.

The number of competing goals in the treatment planning for *intensity modulated radiation therapy* (IMRT) is quite large (about 10-25). Exploring the solutions to the MCO problem involves a high-dimensional Pareto boundary [24]. Approximating this set is computationally expensive, and experience shows that there are many tradeoffs that are either clinically unacceptable or irrelevant. Thus a significant part of the computational effort is wasted on identifying useless solutions.

The complexity of the approximated high-dimensional Pareto boundary (see e.g. [23]) makes it difficult for the decision maker to navigate. An alternative approach would be to first identify the most important aspects of the IMRT plan and make sure that these are met by the solution; compromises, if necessary, would only be acceptable for subsidiary goals.

Toward this end, we want to incorporate a priori knowledge about priorities into the treatment plan optimization process. This information translates naturally into *Lexicographic Optimization* (LO) problems, which constitute a special class of MCO problems.

Classical LO methods minimize the objective functions sequentially, starting with the most important one and proceeding according to the lexicographic

order of the objectives. Here, the optimal value found for each objective is added as a constraint for subsequent optimizations.

These constraints can be quite rigorous, however, depending on the correlation of the corresponding functions and the decision maker might not always want to sacrifice so much optimizing freedom for the sake of a single goal. We therefore propose the more intuitive approach of gathering goals in groups of descending importance, as in [16], and formulating objective functions as weighted sums of the functions contained in these groups. This reduces the number of objective functions to be minimized, which leads to fewer optimization problems to be solved and thus lowers the computational effort.

Typically, the subsequent objectives are in conflict with each other, which implies that the minimum of one objective function is a relatively bad starting point for subsequent optimization levels.

To address this issue, we propose to combine the superiorization methodology with an MCO algorithm. Superiorization can speed up convergence by steering the MCO algorithm towards solutions that are better suited to minimizing subsequent objective functions. In the best case, one could even find the minimizer of all subsequent objective functions before reaching the last optimization level. Here, subsequent minimization becomes unnecessary, and one need only verify that the solution is indeed the minimizer of all subsequent objective functions under the given constraints.

The superiorization methodology [5, 18, 14, 17, 9] was recently developed as a framework for algorithms that lie conceptually between feasibility-seeking and optimization algorithms. It is a heuristic tool that does not guarantee to find the optimum value of a given functional, rather it obtains a solution that is superior (with respect to a given objective function) to the solution achieved by a classical feasibility seeking algorithmic operator.

The main advantage of an algorithm that uses superiorization as the driving tool (as opposed to optimization) is that it requires less computational resources - given a good choice of parameters - while providing comparable solutions, from the point of view of real-world applications, to those that one would get with algorithms that use optimization.

Within this framework, we wish to include the class of projection methods, which are applied successfully in many real-world feasibility problems, see e.g., [1]. This class has witnessed great progress in recent years and its

member algorithms have been applied with success to problems in sensor networks, image reconstruction, image processing, IMRT planning, resolution enhancement and in many others; see [18, 4, 9]. Apart from theoretical interest, the main advantage of projection methods is their computational efficiency. They commonly have the ability to handle huge-size problems of dimensions in which other, more sophisticated methods cease to be efficient.

In this paper we propose to apply the superiorization methodology with projection methods to improve the speed of convergence of lexicographic optimization methods. We demonstrate the plausibility of using superiorization with a simple example with two optimization variables. We also present results from calculations on four real head neck cases in IMRT for two different choices of superiorization parameter sets suited to yield fast convergence for each case individually or robust behavior for all four cases.

The paper is organized as follows. In Section 2 we present some preliminaries, definitions and algorithms that will be needed in the sequel. Later, in Section 3 our general scheme is illustrated on a 2D example and we present the results of our calculations on the aforementioned IMRT head neck cases. Finally we discuss our findings and conclusions in Sections 4 and 5.

2 Methods

2.1 Problems, algorithms and theory

Let us start by giving a definition of what is a multicriteria optimization problem.

Definition 2.1 *Given a mapping $F = (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for all $i \in I = \{1, \dots, m\}$ and let $\{g_j : \mathbb{R}^n \rightarrow \mathbb{R}\}_{j \in J}$ for $J = \{1, \dots, s\}$. The **Multicriteria Optimization (MCO)** problem is the following.*

$$\begin{aligned} & \text{“min” } F(x) \\ & \text{such that } x \in \bigcap_{j \in J} \Omega_j \end{aligned} \tag{2.1}$$

where for all $j \in J$, $\Omega_j := \{x \in \mathbb{R}^n \mid g_j(x) \leq 0\}$.

2.1.1 Lexicographic optimization

In this work we focus on a particular MCO method which is called *lexicographic optimization* (LO). For LO we define an order for the objectives according to their importance and minimize them in that order under the given constraints. After having minimized one objective, we impose an additional constraint which ensures that the objective function value of subsequent solutions do not deviate from the previously found minimum by more than a user-defined value.

In the context of IMRT we consider a more general setting for the formulation of the optimization problems. Here, we gather the objective functions f_1, \dots, f_m in M priority groups which are ranked from most to least important (compare e.g., [16]). Denote by $I_\mu \subseteq I$ an ordered subset of objective function indices such that $\{I_\mu\}_{\mu=1}^M$ is a *partition* of I , meaning that

$$\bigcup_{\mu=1}^M I_\mu = I \text{ and } I_\mu \cap I_\gamma = \emptyset \text{ for } 1 \leq \mu \neq \gamma \leq M. \quad (2.2)$$

It is clear that if for all $\mu = 1, \dots, M$ the set I_μ is a singleton then this reformulation reduces to the classical lexicographic optimization problem.

Now we define

$$\phi_\mu(x) := \sum_{i \in I_\mu} w_i f_i(x) \quad (2.3)$$

where for each $i \in I_\mu$, w_i is the weight of the i -th objective function f_i . So ϕ_μ gathers the objective functions whose indices are elements of I_μ . At optimization level μ ($\leq M$) we thus solve the following problem.

$$\begin{aligned} & \min \phi_\mu(x) \\ & \text{such that } x \in \bigcap_{j \in J} \Omega_j \\ & \phi_\gamma(x) \leq \phi_\gamma^* + \delta_\gamma \quad \gamma = 1, \dots, \mu - 1 \end{aligned} \quad (2.4)$$

where ϕ_γ^* are the minimum values found for each ϕ_γ during the previous optimization levels, $\delta_\gamma \geq 0$ are small user-chosen constants for $\gamma = 1, \dots, \mu - 1$ and Ω_j are as in (2.1).

Let us illustrate the above for a two-stage LO also known as *bi-level optimization*. Assume that $I_1 = \{1\}$, $I_2 = \{2\}$, and $w_1 = w_2 = 1$; Thus

$\phi_1 = f_1$ and $\phi_2 = f_2$. At the first optimization level we solve the following single-objective optimization problem.

$$\begin{aligned} \min \quad & f_1(x) \\ \text{such that } & x \in \bigcap_{j \in J} \Omega_j. \end{aligned} \tag{2.5}$$

After we have solved (2.5) and obtained f_1^* ($= \phi_1^*$), we continue with the second optimization level, where our problem formulates as

$$\begin{aligned} \min \quad & f_2(x) \\ \text{such that } & x \in \bigcap_{j \in J} \Omega_j \\ & x \in \Omega^{\delta,1} \end{aligned} \tag{2.6}$$

where $\Omega^{\delta,1} := \{x \in \mathbb{R}^n \mid f_1(x) \leq f_1^* + \delta_1\}$ for some small user-chosen constant $\delta_1 \geq 0$.

In the above, the choice of δ_1 and in general δ_μ for $\mu = 1, \dots, M$, defines in a fixed way how far from previously found optima of objective functions of higher priority the algorithm is allowed to deviate in the current optimization level. A different approach has been taken by [25] which allows the decision maker to choose the tradeoff between two subsequent (in terms of their lexicographic order) objective functions in an interactive way.

Observe that if the constraint set of (2.5)

$$\bigcap_{j \in J} \Omega_j \tag{2.7}$$

(or of (2.6) with $\delta_1 = 0$) contains only one element, then this is the optimal solution of the bi-level optimization problem and it is found already at the first optimization level. Otherwise, in case of multiple feasible points, a feasibility seeking algorithm for finding an element of (2.5) might end up with a point which is relatively far from the optimal solution of the second optimization level, this is due to the fact that the second objective function has not been taken into account yet. This effect is likely to occur when the objective functions of the subsequent optimization levels represent conflicting goals as commonly happens in cancer therapy.

Motivated by these difficulties we propose to steer the algorithm towards solutions that provide good starting points for subsequent optimization levels by applying the superiorization methodology together with projection methods. Before we can do so we need to transform the optimization problem such that we can use projection methods to solve it.

2.1.2 Level set scheme

For the aforementioned purpose we use the *level set scheme* [12] which is a new projection-based scheme for solving convex optimization problems. It transforms a convex optimization problem into a sequence of auxiliary feasibility problems by iteratively constraining the objective function from above until the feasibility problem is inconsistent. To illustrate the idea let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex and continuously differentiable function and $\Xi \subseteq \mathbb{R}^n$ nonempty, closed and convex set. Following the level set scheme we can reformulate the problem

$$\begin{aligned} \min \quad & \phi(x) \\ \text{such that } & x \in \Xi \end{aligned} \tag{2.8}$$

in the following form

$$\begin{aligned} \min \quad & t \\ \text{such that } & \phi(x) \leq t \\ & x \in \Xi. \end{aligned} \tag{2.9}$$

Now let $\{\varepsilon_k\}_{k=0}^\infty$ be some user chosen positive sequence, for example $\varepsilon_k \equiv \varepsilon = 0.1$ or $\varepsilon_k = 0.1 |\phi(x^k)|$.

In the initial step of the level set scheme we choose any projection method to find a feasible point $x^0 \in \Xi$, set $t_0 = \phi(x^0) - \varepsilon_0$ and $k = 0$. Now at the iterative step $k > 0$, when we are given the current point x^{k-1} , we try to solve the following *convex feasibility problem* (CFP)

$$\begin{aligned} \text{Find a point } & x^k \in \mathbb{R}^n \\ \text{such that } & \phi(x^k) \leq t_k \\ & x^k \in \Xi. \end{aligned} \tag{2.10}$$

If there exists a feasible solution, set $t_{k+1} = \phi(x^k) - \varepsilon_k$ and continue, else there exists no feasible solution, and x^k is an ε_k -optimal solution.

2.1.3 Projection methods

In this work, we use *projection methods* to solve the auxiliary convex feasibility problems generated by the level set scheme. Projection methods are iterative algorithms that use projections onto sets. They rely on the general principle that when a family of (usually closed and convex) sets is present, projections onto the given individual sets are easier to perform than projections onto other sets (intersections, image sets under some transformation, etc.) that are derived from the given individual sets. Projection methods come in various algorithmic structures, some of which are particularly suitable for parallel computing, and they demonstrate nice convergence properties, for example bounded perturbations resilience. This fact allows us to use them to solve the CFPs resulting from the level set scheme and to incorporate superiorization, which is presented in the following paragraph, while retaining convergence to the optimal point.

Consider the CFP:

$$\text{Find a point } x \in \bigcap_{j \in J} \Omega_j \quad (2.11)$$

with Ω_j defined as in (2.1). The particular method we use in this paper is the *simultaneous subgradient projections method* which is defined as follows. Let $x^0 \in \mathbb{R}^n$ be an arbitrary starting point. Given the current iterate x^k , calculate the next iterate x^{k+1} via

$$x^{k+1} := x^k - \lambda_k \sum_{j \in J, g_j(x^k) > 0} w_j \frac{g_j(x^k)}{\|\xi^k\|^2} \xi^k \quad (2.12)$$

where $\xi^k \in \partial g_j(x^k)$ (subgradient of g_j at x^k), $w_j > 0$ are *weights* and $\lambda_k \in [\epsilon_1, 2 - \epsilon_2]$ (relaxation parameters) for arbitrary $\epsilon_1, \epsilon_2 > 0$.

Simultaneous projection methods are also referred to as “parallel” methods. In this case, in order to evaluate the next iterate, all (or a block of more than one) constraints are taken into account. In case of a system of linear equalities, this method is known as Cimmino’s method [7]. See [15, 10] for more details.

Definition 2.2 *Given a problem \mathbb{P} , an algorithmic operator $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be **bounded perturbations resilient** if the following is true.*

If the sequence $\{x^k\}_{k=0}^\infty$, generated by $x^{k+1} = \mathbf{T}(x^k)$, for all $k \geq 0$, converges to a solution of \mathbb{P} , then any sequence $\{y^k\}_{k=0}^\infty$ of points in \mathbb{R}^n generated by $y^{k+1} = \mathbf{T}(y^k + \beta_k v^k)$, for all $k \geq 0$, also converges to a solution of \mathbb{P} provided that, for all $k \geq 0$, $\beta_k v^k$ are bounded perturbations, meaning that $\beta_k \geq 0$ for all $k \geq 0$ such that $\sum_{k=0}^\infty \beta_k < \infty$ and the sequence $\{v^k\}_{k=0}^\infty$ is bounded.

2.1.4 Superiorization methodology

The *superiorization methodology* of [14] was recently developed as a framework for algorithms that lie conceptually between feasibility-seeking and optimization algorithms. It is designed to find a solution to a CFP which is superior with respect to a given objective function ψ , meaning with a value of ψ at least as low, but possibly lower, compared to the solution obtained by a classical feasibility seeking algorithmic operator. The state of current research on superiorization can best be appreciated from the “Superiorization and Perturbation Resilience of Algorithms: A Bibliography compiled and continuously updated by Yair Censor” which is at: <http://math.haifa.ac.il/yair/bib-superiorization-censor.html>. In particular, [13] and [3] are recent reviews of interest.

Definition 2.3 Given a function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ and a point $z \in \mathbb{R}^n$, we say that a vector $d \in \mathbb{R}^n$ is **nonascending** for ψ at z if and only if $\|d\| \leq 1$ and there is a $\zeta > 0$ such that for all $\lambda \in [0, \zeta]$ we have $\psi(z + \lambda d) \leq \psi(z)$.

Consider again the convex feasibility problem:

$$\text{Find a point } x \in \Xi \tag{2.13}$$

where $\Xi \subseteq \mathbb{R}^n$ is a closed and convex set and let \mathbf{T} be any feasibility-seeking algorithmic operator, that defines the basic algorithm $x^{k+1} = \mathbf{T}(x^k)$ to solve the convex feasibility problem.

Let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a given convex and continuously differentiable function. The *superiorized* (with respect to ψ) version of \mathbf{T} is

$$y^{k+1} = \mathbf{T}(y^k + \beta_k d^k) \tag{2.14}$$

where d^k is nonascending for ψ at y^k .

One option for d^k , as it is used in [14], is

$$d^k = \begin{cases} -\frac{\nabla\psi(y^k)}{\|\nabla\psi(y^k)\|} & \text{if } \nabla\psi(y^k) \neq 0 \\ 0 & \nabla\psi(y^k) = 0. \end{cases} \quad (2.15)$$

If the solution set of the feasibility problem contains several elements, the superiorized version of \mathbf{T} obtains a solution that is superior with respect to ψ to the solution achieved by the basic algorithmic operator \mathbf{T} .

2.1.5 Main result

We now combine all of the mentioned concepts to derive our approach. We consider the multicriteria optimization problem (2.1) and define the priority groups I_μ and weights $\{w_i\}_{i \in I_\mu}$, $\mu = 1, \dots, M$. We then transform it into a LO problem like (2.4). Now we apply the level set scheme and consequently solve the following optimization problem at every optimization level μ :

$$\begin{aligned} \min \quad & t^{(\mu)} \\ \text{such that } & x \in \Omega^{t^{(\mu)}} \\ & x \in \bigcap_{j \in J} \Omega_j \\ & x \in \bigcap_{\gamma=1}^{\mu-1} \Omega^{\delta, \gamma} \end{aligned} \quad (2.16)$$

where

$$\begin{aligned} \Omega^{t^{(\mu)}} &:= \{x \in \mathbb{R}^n \mid \phi_\mu(x) \leq t^{(\mu)}\} \\ \Omega_j &:= \{x \in \mathbb{R}^n \mid g_j(x) \leq 0\} \\ \Omega^{\delta, \gamma} &:= \{x \in \mathbb{R}^n \mid \phi_\gamma(x) \leq \phi_\gamma^*(x) + \delta_\gamma\}. \end{aligned}$$

To do that, we transform (2.16) into a sequence of auxiliary CFPs of the form:

$$\begin{aligned}
& \text{Find a point } x^k \in \mathbb{R}^n \\
& \text{such that } x^k \in \Omega_k^{t_k^{(\mu)}} \\
& x^k \in \bigcap_{j \in J} \Omega_j \\
& x^k \in \bigcap_{\gamma=1}^{\mu-1} \Omega^{\delta, \gamma}
\end{aligned} \tag{2.17}$$

for decreasing sequences of $\{t_k^{(\mu)}\}$. We solve the auxiliary CFPs by using the simultaneous subgradient projections method as the basic algorithm **T**.

Let $\mathcal{S} \subseteq \bigcup_{\gamma=\mu+1}^M I_\gamma$ be a subset of indices contained in the subsequent priority groups $I_{\mu+1}, \dots, I_M$ and let \hat{w}_i be weights to the corresponding objective functions.

Then, after having successfully solved K auxiliary CFPs we include superiorization with respect to $\psi(x) := \sum_{i \in \mathcal{S}} \hat{w}_i f_i$ by defining d^k as in (2.15) with x^k and setting

$$x^{k+1} = x^k + \beta_k d^k \tag{2.18}$$

where $\beta_k \geq 0$ for all $k \geq 0$ such that $\sum_{k=0}^{\infty} \beta_k < \infty$ and $\psi(x^k + \beta_k d^k) \leq \psi(x^k)$, meaning that d^k is a direction of nonascend for ψ at x^k .

In the following we refer to this method as *superiorized lexicographic optimization* (SLO).

2.2 Application to IMRT

The IMRT planning problem is to find a treatment plan, i.e. energy fluence intensities, which results in a dose distribution in the patient's body that irradiates the tumor as homogeneously as possible while sparing critical healthy organs.

The dose (measured in Gy) delivered to the patient's body is evaluated according to clinical goals which are translated into the functions given below. The clinical goals are ordered according to their importance and can

also be grouped together which reduces the number of optimization levels and therefore accelerates the optimization process. For our calculations on four head neck cases we chose one group of goals as constraints and three subsequent priority groups.

2.2.1 Problem formulation

The volume of the patient's body is discretized into a three-dimensional grid of voxels. The voxels are assigned to different planning structures, e.g. organs or tumor tissue. Let \mathbf{P} denote the so-called dose matrix that maps a vector x of fluence intensities to the dose $d = \mathbf{P}x$ received by the individual voxels. Thus a planning structure \mathcal{O} is essentially a subset of indices of the vector d .

To evaluate the dose received by different planning structures we used lower and upper tail penalty functions as well as mean upper tail penalty functions. The lower tail penalty function of the dose d for a planning structure \mathcal{O} penalizes dose values below a given threshold L . It is given by

$$f_{\mathcal{O},low}(d) := (1/|\mathcal{O}|) \sum_{i \in \mathcal{O}} (\max\{L - d_i, 0\})^2 \quad L \in \mathbb{R}. \quad (2.19)$$

The upper tail penalty function is

$$f_{\mathcal{O},up}(d) := (1/|\mathcal{O}|) \sum_{i \in \mathcal{O}} (\max\{d_i - U, 0\})^2 \quad U \in \mathbb{R} \quad (2.20)$$

and is used to penalize dose values exceeding a given threshold U . The mean upper tail function is defined as

$$f_{\mathcal{O},mean}(d) := \left(\max \left\{ \left((1/|\mathcal{O}|) \sum_{i \in \mathcal{O}} d_i \right) - M, 0 \right\} \right)^2 \quad M \in \mathbb{R}. \quad (2.21)$$

This function is used to keep the mean dose values below a given threshold M , which is a less strict way of avoiding overdosage than using the upper tail penalty function.

Table 1 shows our problem formulation for the first head neck cases according to [21]. The other cases were treated similarly, depending on the individual geometry of the tumor tissue. To gain an impression of the geometry of some of the planning structures involved see Figure 2. Note that the function values of the evaluation functions listed as constraints were

required to be 0 which translates to the dose prescription of the corresponding planning structure being met by the dose distribution.

The weights for the evaluation functions were chosen empirically for each case individually such that the optimization would yield a clinically acceptable treatment plan.

Non-tumor Tissue 1 denotes the three-dimensional margin of 1 cm around the lymphatic drainage pathways (which provide the largest of the target volumes and contain both PTV 60 and PTV 70). Non-tumor Tissue 2 is the tissue that consists of everything but the lymphatic drainage pathways and Non-tumor Tissue 1.

These structures are introduced to reduce the radiation to non-tumor tissue, especially to areas which are not covered by any of the organs at risk. Additionally, it offers an incentive to the optimization algorithm to reduce the dose values of the organs at risk that lie outside of the lymphatic drainage pathways even below the prescribed maximum or maximum mean dose values.

In optimization level μ we minimize the objective function ϕ_μ which is a weighted sum of the functions belonging to priority group I_μ .

Table 1: Evaluation functions of the structures under consideration gathered in different priority groups.

planning structure \mathcal{O}	function	parameters	priority group
PTV 70	lower tail penalty	$L = 66.5$	constraint
PTV 70	upper tail penalty	$U = 80.5$	constraint
PTV 70	mean upper tail penalty	$M = 73.5$	constraint
Myelon	upper tail penalty	$U = 45$	constraint
PTV 60	lower tail penalty	$L = 57.5$	I_1
Eye right	upper tail penalty	$U = 45$	I_1
Eye left	upper tail penalty	$U = 45$	I_1
Optic nerve right	upper tail penalty	$U = 50$	I_1
Optic nerve left	upper tail penalty	$U = 50$	I_1
Parotid left	mean upper tail penalty	$M = 26$	I_1
Lymphatic drainage	lower tail penalty	$L = 47.5$	I_2
Parotid right	mean upper tail penalty	$M = 26$	I_2
Brain	upper tail penalty	$U = 60$	I_2
Brainstem	upper tail penalty	$U = 52$	I_2
Larynx	mean upper tail penalty	$M = 45$	I_2
Esophagus	mean upper tail penalty	$M = 45$	I_2
Oral cavity	mean upper tail penalty	$M = 40$	I_2
Non-tumor Tissue 2	mean upper tail penalty	$M = 5$	I_2
Plexus right	upper tail penalty	$U = 63$	I_3
Plexus left	upper tail penalty	$U = 63$	I_3
Mandible	upper tail penalty	$U = 75$	I_3
Inner ear right	upper tail penalty	$U = 45$	I_3
Inner ear left	upper tail penalty	$U = 45$	I_3
Lips	upper tail penalty	$U = 30$	I_3
Non-tumor Tissue 1	mean upper tail penalty	$M = 20$	I_3

2.2.2 Implementation

In this section we present pseudo codes of the algorithms we implemented to solve the IMRT optimization problems. Algorithm 1 implements the lexicographic optimization problem turned into a sequence of convex feasibility problems and Algorithm 2 implements the superiorization methodology.

In the following paragraph we give some details about the submethods mentioned in the codes.

findFeasibleSolution($t^{(\mu)}$) tries to solve the CFP (2.17) with $t_k^{(\mu)} = t^{(\mu)}$ within n_{max} simultaneous projections. If this is successful, it returns $[x, \text{true}]$ with $\phi_\mu(\mathbf{P}x) \leq t_k^{(\mu)}$ and x feasible. If this is not successful, the method returns $[x, \text{false}]$ with $\phi_\mu(\mathbf{P}x) \leq \tilde{t}^{(\mu)}$, where $\tilde{t}^{(\mu)} > t_k^{(\mu)}$ and x feasible. $\tilde{t}^{(\mu)}$ is the smallest upper bound for ϕ_μ that the algorithm was able to find a feasible solution for.

reduce($\phi_\mu(\mathbf{P}x)$) returns a new upper bound $t^{(\mu)}$ for ϕ_μ that is smaller than $\phi_\mu(\mathbf{P}x)$.

addConstraint(ϕ_μ^*) adds $\phi_\mu(x) \leq \phi_\mu^* + \delta_\mu$ to the set of constraints.

supDirection(x) returns the direction of superiorization at x . In our calculations we used superiorization with respect to the objective function $\phi_{\mu+1}$ of the following optimization level.

Algorithm 1 Lexicographic level set scheme

```
1:  $\mu = 1$ ;  
2: while  $\mu \leq M$  do  
3:   if  $\mu == 1$  then  
4:      $t^{(\mu)} = \infty$   
5:   else  
6:      $t^{(\mu)} = \text{reduce}(\phi_\mu(\mathbf{P}x))$   
7:   end if  
8:    $\text{levelsSolved} = \text{false}$   
9:    $\text{counter} = 0$   
10:  while  $\neg \text{levelsSolved}$  do  
11:     $[\mathbf{x}, \text{feasibleSolutionFound}] = \text{findFeasibleSolution}(t^{(\mu)})$   
12:     $\text{counter}++$   
13:    if  $\text{feasibleSolutionFound}$  then  
14:      if  $t^{(\mu)} == t_{\min}$  then  
15:         $\text{levelsSolved} = \text{true}$   
16:         $\phi_\mu^* = \phi_\mu(\mathbf{P}x)$   
17:         $\text{addConstraint}(\phi_\mu^*)$   
18:         $\mu++$   
19:      else  
20:        if  $\mu < M \ \&\& \ \text{counter} \pmod{K} == 0$  then  
21:           $\mathbf{x} = \text{superiorize}()$   
22:        end if  
23:         $t^{(\mu)} = \text{reduce}(\phi_\mu(\mathbf{P}x))$   
24:      end if  
25:    else  
26:       $\text{levelsSolved} = \text{true}$   
27:       $\phi_\mu^* = \phi_\mu(\mathbf{P}x)$   
28:       $\text{addconstraint}(\phi_\mu^*)$   
29:       $\mu++$   
30:    end if  
31:  end while  
32: end while
```

Algorithm 2 Superiorize

```
1: coefficientBigEnough = true
2: base  $\in (0, 1)$ 
3:  $\lambda = 0$ 
4: while  $\lambda < \Lambda$  && coefficientBigEnough do
5:   direction = supDirection( $x$ )
6:   loop = true
7:   exponent = 1
8:   while loop && coefficientBigEnough do
9:     coefficient = baseexponent
10:     $x_{sup} = x + \text{coefficient} \cdot \text{direction}$ 
11:    if  $x_{sup} \geq 0$  &&  $\phi_{\mu+1}(\mathbf{P}x_{sup}) \leq \phi_{\mu+1}(\mathbf{P}x)$  then
12:       $x = x_{sup}$ 
13:      loop = false
14:       $\lambda++$ 
15:    else
16:      if coefficient > minStepsize then
17:        exponent++
18:      else
19:        coefficientBigEnough = false
20:      end if
21:    end if
22:  end while
23: end while
```

3 Numerical illustration

In the first part of this section we demonstrate the plausibility of using the superiorized lexicographic optimization methodology on a simple example with two optimization variables which is taken from Stanimirović [19]. In the second part we present the performance of our method on four real IMRT head neck cases.

3.1 2D example

To demonstrate the mechanism of superiorization and how it can be useful for lexicographic optimization we consider the following example.

Example 3.1 *Solve the following multicriteria optimization problem*

$$\begin{aligned}
\min \quad & \Phi(x) = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} -8x_1 - 12x_2 \\ -14x_1 - 10x_2 \\ -x_1 - x_2 \end{pmatrix} \\
\text{such that} \quad & 2x_1 + x_2 - 150 \leq 0 \\
& 2x_1 + 3x_2 - 300 \leq 0 \\
& 4x_1 + 3x_2 - 360 \leq 0 \\
& -x_1 - 2x_2 + 120 \leq 0 \\
& -x_1 \leq 0 \\
& -x_2 \leq 0
\end{aligned} \tag{3.1}$$

where the functions ϕ_1 , ϕ_2 and ϕ_3 are given in lexicographic order.

The optimal solution $x^* = (x_1^*, x_2^*)$ is $(30, 80)$ with $\Phi(x^*) = (\phi_1(x^*), \phi_2(x^*), \phi_3(x^*)) = (-1200, -1220, -110)$.

We stopped the optimization process at the iterate x^k if $\|\Phi(x^k) - \Phi(x^*)\| \leq 10^{-2}$. In the following, points are given with precision 10^{-2} . In this example calculating the gradients and projections is computationally very cheap, but for IMRT cases gradient- and dose-evaluations consume most of the computation time, so we compare the number of projections and gradient evaluations.

Figure 1 shows the feasible region and the trajectories of the iterates of the classical and the superiorized LO for the starting point $x^0 = (0, 47.5)$.

Both methods first seek feasibility and find the point $(5, 57.5)$. In optimization level 1, the classical level set scheme minimizes with respect to ϕ_1 only and ends up at the solution $x^{(1)} = (23.08, 84.625)$ of the first optimization level with $\phi_1^* = \phi_1(x^{(1)}) = -1200$ using 327 projections and 327 gradient calculations. Additionally, we need another 260 projections and 518 gradient evaluations to find out that we cannot improve ϕ_1 any further while staying feasible. In optimization level 2, we want to preserve $\phi_1(x^k) = \phi_1^*$ for all subsequent iterates x^k and minimize ϕ_2 under this additional condition. We arrive at the solution $x^{(2)} = (30, 80)$ of optimization level 2 after a total of 4743 projections and 9107 gradient calculations. We cannot improve any further with respect to ϕ_3 and have found the optimal solution.

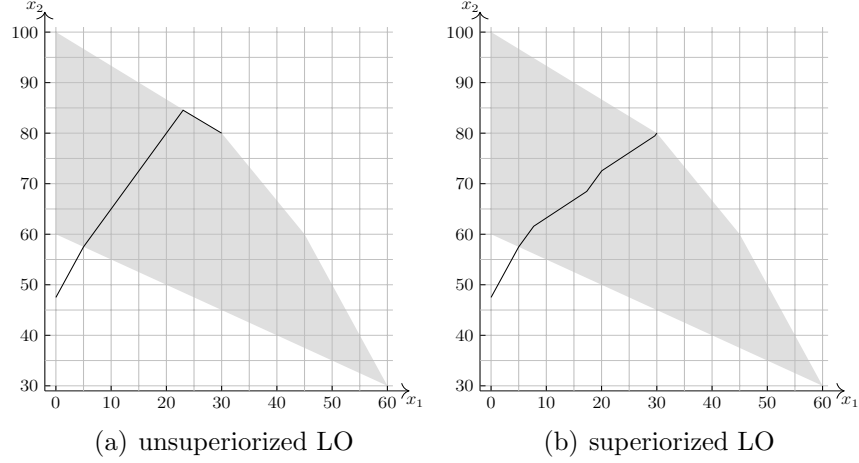


Figure 1: Trajectories for unsuperiorized and superiorized LO. The set of feasible solutions is shown in light gray.

The superiorized level set scheme minimizes with respect to ϕ_1 in optimization level 1 but uses superiorization with respect to ϕ_2 . It arrives at the solution $x^{(1)} = (30, 80)$ of optimization level 1 after 108 projections and 144 gradient calculations, 34 of which were needed for the superiorization. Using the superiorized LO we get already sufficiently close to the optimum in optimization level 1.

3.2 IMRT

From the formulae (2.19)–(2.21) of the utilized functions it is clear that the minimal value we can hope to achieve for any of the ϕ_μ is 0. Thus we considered an optimization level μ to be solved, if $\phi_\mu(\mathbf{P}x) \leq t_{min} := 10^{-8}$ given that x is feasible.

What takes most of the computational time with IMRT cases are multiplications of the dose matrix \mathbf{P} with the fluence vector x which occur when we evaluate the dose or calculate the objective function gradient for a given x . The number of these multiplications is what we will use as units to measure progress, i.e., reduction of objective function values, of the classical and the superiorized LO.

In Algorithm 1 and 2 in Subsection 2.2.2 we find parameters K and Λ which

determine the behavior of the method. K is the number of CFPs that have to be successfully solved before superiorization is applied and Λ is the maximum number of superiorization steps that are taken. If the superiorization step size is smaller than a predefined threshold or the maximum number of steps is reached, the superiorization stops and we solve the next CFP.

In Tables 2 and 3 we first present results we achieved by trying different parameter sets (K, Λ) and then picking for each case individually the set that resulted in the fastest convergence of the algorithm measured in the number of multiplications. We refer to the sets of parameters picked according to this strategy as $(\hat{K}, \hat{\Lambda})$. All values presented in the tables in this section are rounded to 10^{-4} .

The results presented in Table 2 show that fast SLO results in equal or lower objective values than unsuperiorized LO within less multiplications. In Table 3 we see that fast SLO produces solutions of the single optimization levels with potentially lower objective values of the subsequent objective functions than unsuperiorized LO. Fast SLO thus offers a better starting point to the single optimization levels than unsuperiorized LO which reduces the overall number of multiplications.

$\phi_\mu^* = \phi_\mu(x^{(\mu)})$ denotes the value of ϕ_μ at the solution $x^{(\mu)}$ of optimization level μ obtained using the classical LO, $\hat{\phi}_\mu^* = \phi_\mu(\hat{x}^{(\mu)})$ denotes the value of ϕ_μ at the solution $\hat{x}^{(\mu)}$ of optimization level μ obtained using the superiorized LO with $(\hat{K}, \hat{\Lambda})$.

N_j^k is the number of multiplications needed by the classical LO to solve optimization level $j+1$ to level k . $N := N_0^M = \sum_{j=0}^{M-1} N_j^{j+1}$ is the total number of multiplications. Analogously, \hat{N}_j^k denotes the number of multiplications needed by the superiorized LO with $(\hat{K}, \hat{\Lambda})$ to solve optimization level $j+1$ to level k and $\hat{N} := \sum_{j=0}^{M-1} \hat{N}_j^{j+1}$.

For our calculations we chose $\delta_\mu = 0.1\phi_\mu^*$ for the classical LO and $\delta_\mu = 0.1\hat{\phi}_\mu^*$ for the superiorized LO.

Table 2: Optimal objective value and total multiplication number ratios for $(\hat{K}, \hat{\Lambda})$. Fast SLO achieves lower or equal objective values than unsuperiorized LO within fewer multiplications. The notation 1^\diamond means that we solved the optimization level, meaning that the objective function value has dropped below 10^{-8} .

	$\hat{\phi}_1^*/\phi_1^*$	$\hat{\phi}_2^*/\phi_2^*$	$\hat{\phi}_3^*/\phi_3^*$	\hat{N}/N
case 1	1^\diamond	0.9598	0.9923	0.8998
case 2	1^\diamond	0.9743	0.9960	0.6612
case 3	1^\diamond	0.9996	1.0000	0.9210
case 4	1^\diamond	0.9708	0.9941	0.8959

Table 3: Multiplication numbers per optimization level and objective values of starting points for $(\hat{K}, \hat{\Lambda})$. Fast SLO produces solutions of the single optimization levels with potentially lower objective values of the subsequent objective functions than unsuperiorized LO.

	\hat{N}_0^1/N_0^1	\hat{N}_1^2/N_1^2	\hat{N}_2^3/N_2^3	$\phi_2(\hat{x}^{(1)})/\phi_2(x^{(1)})$	$\phi_3(\hat{x}^{(2)})/\phi_3(x^{(2)})$
case 1	1.6521	0.5669	0.9895	0.3741	0.9856
case 2	1.4066	0.3588	1.0065	0.3682	1.0008
case 3	1.0403	0.8594	1.0000	0.5625	1.0075
case 4	0.3818	1.0536	0.8591	0.8196	0.9939

It is not clear a priori how to choose values for K and Λ for a given problem. Therefore we aimed to find a robust set of parameters that yields good results for all IMRT cases and could possibly be applied for other IMRT head neck cases as well. We refer to this set of parameters as $(\check{K}, \check{\Lambda})$. The results for the robust parameter set are presented in Table 4 and 5.

The results in Table 4 show that robust SLO achieves comparable objective values to unsuperiorized LO within fewer multiplications. In Table 5 we see that robust SLO, too, produces solutions of the single optimization levels with potentially lower objective values of the subsequent objective functions than unsuperiorized LO. Like fast SLO, robust SLO also offers a better starting point to the single optimization levels than unsuperiorized LO which again reduces the overall number of multiplications.

Following the previous notation $\check{\phi}_\mu^* = \phi_\mu(\check{x}^{(\mu)})$ is the value of ϕ_μ at the solution $\check{x}^{(\mu)}$ of optimization level μ obtained using the superiorized LO with $(\check{K}, \check{\Lambda})$.

\check{N}_j^k denotes the number of multiplications needed by the superiorized LO with $(\check{K}, \check{\Lambda})$ to solve optimization level $j + 1$ to level k and $\check{N} := \sum_{j=0}^{M-1} \check{N}_j^{j+1}$.

Table 4: Optimal objective value and total multiplication number ratios for $(\check{K}, \check{\Lambda})$. Robust SLO achieves comparable objective values to unsuperiorized LO within fewer multiplications. The notation 1^\diamond means that we solved the optimization level, meaning that the objective function value has dropped below 10^{-8} .

	$\check{\phi}_1^*/\phi_1^*$	$\check{\phi}_2^*/\phi_2^*$	$\check{\phi}_3^*/\phi_3^*$	\check{N}/N
case 1	1^\diamond	1.0046	1.0002	0.9177
case 2	1^\diamond	1.0020	1.0043	0.8877
case 3	1^\diamond	0.9829	0.9969	0.9541
case 4	1^\diamond	1.0245	1.0071	0.8859

Table 5: Multiplication numbers per optimization level and objective values of starting points for $(\check{K}, \check{\Lambda})$. Robust SLO produces solutions of the single optimization levels with potentially lower objective values of the subsequent objective functions than unsuperiorized LO.

	\check{N}_0^1/N_0^1	\check{N}_1^2/N_1^2	\check{N}_2^3/N_2^3	$\phi_2(\check{x}^{(1)})/\phi_2(x^{(1)})$	$\phi_3(\check{x}^{(2)})/\phi_3(x^{(2)})$
case 1	1.1825	0.6972	0.9849	0.3916	1.0008
case 2	0.7447	0.8409	0.9999	0.5194	1.0058
case 3	1.1320	0.9029	0.9999	0.5530	1.0007
case 4	1.2453	0.8575	0.8664	0.4959	1.0103

In Figure 2 we present one slice of the CT of case 1 with the dose distribution resulting from the solution of superiorized LO with the robust parameter set $(\check{K}, \check{\Lambda})$. The corresponding DVH is shown in Figure 3.

Both the images and the dose matrix were created using CERR [22]. Note that the dose calculation via the dose matrix, which is used in the optimization algorithm, provides only an approximation to the dose values calculated by

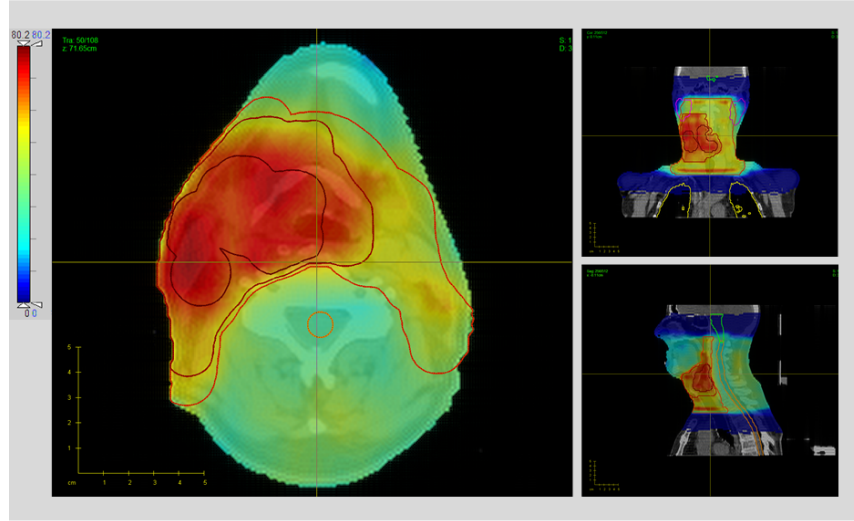


Figure 2: CT image and colorwash of dose resulting from robustly superiorized LO. The contours of the target volumes PTV 70, PTV 60 and the lymphatic drainage pathways (LDP) are shown in dark, medium and bright red. The contours of the left and right parotid are marked in dark and bright magenta. The lungs are contoured in yellow and the myelon is shown in orange. Finally, the brainstem is contoured in bright light green.

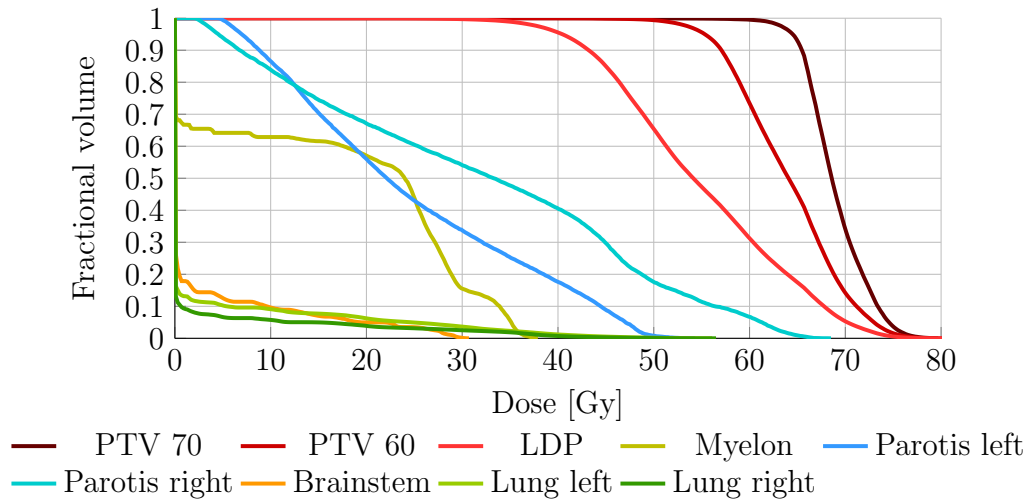


Figure 3: DVHs resulting from solution dose of robustly superiorized LO.

CERR which are depicted in Figure 2 and 3. Thus the DVH curves show e.g. that dose values below 66.5 Gy occur in the PTV70 even though in the optimization the mathematical constraint ensuring that this does not happen is not violated.

In Figure 4 we present the objective values of ϕ_1 , ϕ_2 and ϕ_3 plotted against the number of dosematrix-vector-multiplications for case 1. To show the fundamental behavior of the superiorization method we intentionally did not give any actual numbers in the plots in Figure 4.

In Figure 4(a) we can see how ϕ_2 increases while we minimize ϕ_1 in level 1 using classical LO. We see how the superiorization with respect to ϕ_2 disturbs the minimization of ϕ_1 and leads to N_0^1 being smaller than \hat{N}_0^1 . Figure 4(b) shows that $\hat{x}^{(1)}$ is a better starting point for the minimization of ϕ_2 in level 2 than $x^{(1)}$ because its objective value of ϕ_2 is significantly lower. This results in \hat{N}_1^2 being notably smaller than N_1^2 .

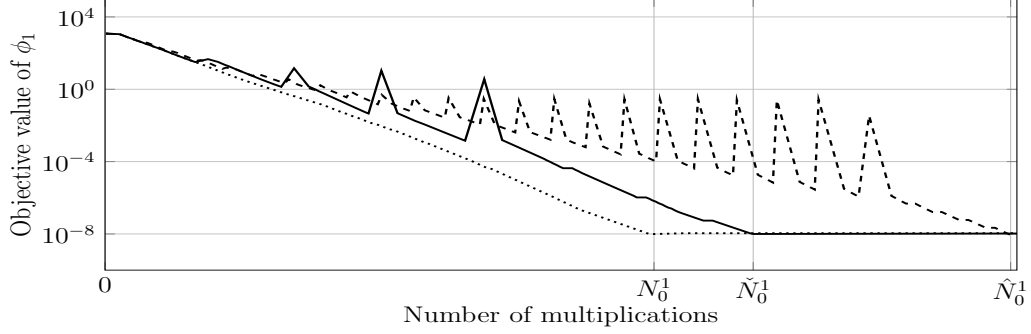
Due to the correlation of the clinical goals there are relatively few auxiliary CFPs to be solved in optimization level 2 for all cases. When superiorization is applied (both with optimal and robust (K, Λ)) we even get so close to the minimum of ϕ_2 at the end of level 1 that for these parameter pairs the CFP in level 2 turns inconsistent within less than \hat{K} or \check{K} reductions of $t_k^{(2)}$. Therefore no superiorization is applied in level 2 and the objective value of ϕ_3 is about the same for fast SLO, robust SLO and unsuperiorized LO.

As we can see from Figure 4(c) superiorization with respect to ϕ_2 during level 1 can in some cases – depending on the correlation of the objective functions – reduce the value of ϕ_3 and \hat{N}_2^3 .

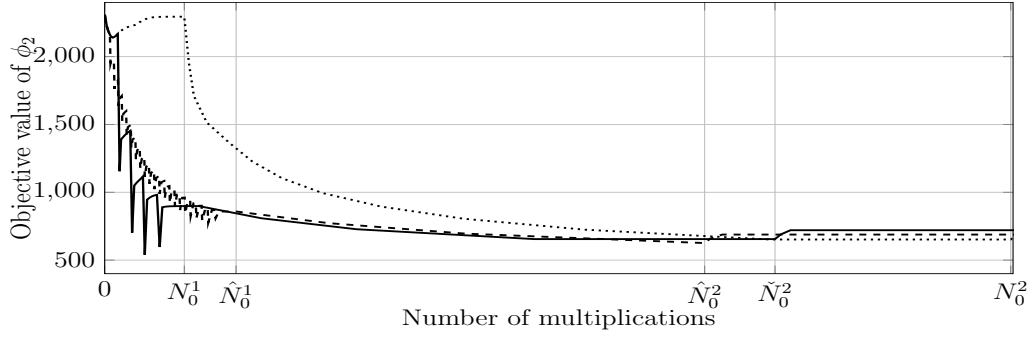
Similar behavior can be observed for the superiorized LO that uses the robust parameter set $(\check{K}, \check{\Lambda})$.

4 Discussion

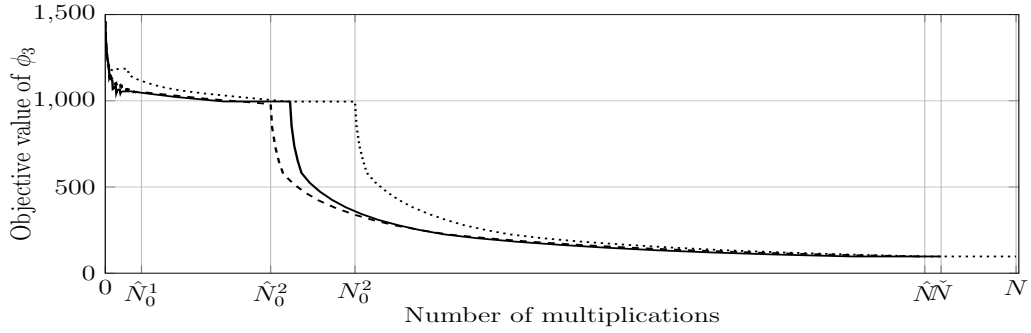
We used LO as an intuitive approach to lower the complexity of decision making in IMRT and to include a priori knowledge about the priorities of the decision maker. We combined this MCO technique with the superiorization methodology to speed up the optimization in the next optimization level. This is not the only application for superiorization. One could for example choose as superiorization function ψ a weighted sum of different evaluation functions than we did, or one could choose ψ as a function that is not included in the lexicographic optimization model at all. In this work we identified the



(a) Values of ϕ_1 for robust SLO, fast SLO and unsuperiorized LO.



(b) Values of ϕ_2 for robust SLO, fast SLO and unsuperiorized LO.



(c) Values of ϕ_3 for robust SLO, fast SLO and unsuperiorized LO.

— robust SLO --- fast SLO unsuperiorized LO

Figure 4: Superiorization perturbs the optimization of ϕ_1 . Superiorization applied in level 1 offers a starting point with significantly lower value of ϕ_2 and speeds up the optimization of ϕ_2 . No superiorization is applied in level 2 and SLO and unsuperiorized LO perform similarly. Due to the speedup in level 2 the overall number of multiplications is lower for SLO.

superiorization parameters manually. This process takes of course too much time to be of any practical use. To be able to find robust parameters that work for many optimization problems in an automated way, one should make sure to identify problems with similar tradeoffs among the objective functions and the superiorization function(s).

An issue with using projection methods to solve CFPs is that in general we do not have a criterion that tells us whether a CFP is inconsistent. Therefore we defined a maximum number n_{max} of attempts, i.e. simultaneous projections, to solve it. If the CFP cannot be solved within n_{max} simultaneous projections, we assume it to be inconsistent. This means that finding out that a CFP actually is inconsistent takes n_{max} projections, which drastically increases the total number of projections.

On the other hand, low values of n_{max} might not suffice to reach the actual minimum of an objective function under the given constraints, but only an approximation thereof.

We decided to stop solving an optimization level μ after encountering the first inconsistent auxiliary CFP (2.17) and continue with level $\mu + 1$, accepting that we might miss the minimal value of ϕ_μ accessible with the chosen value of n_{max} by a certain gap (which is bounded by the reduction strategy of $t^{(\mu)}$, in our case by 10 %).

With this strategy, the computationally expensive process of finding out that a CFP is inconsistent occurs (at most) once per optimization level.

For this reason it is a big advantage to know the objective values of the optimal solution a priori as in Example 3.1. In general we do not have this information but as in our calculations for the IMRT cases, sometimes it is possible to exploit a priori information about the objective functions to define a stopping criterion that is computationally cheap to evaluate.

A question arising in the context of optimization problems is how algorithm 1 performs compared to other solvers. In the following, we will refer to algorithm 1 as (superiorized) projection solver. We implemented the projection solver in Matlab. To keep computation time for algebraic operations comparable we chose to compare our solutions to the solutions of Matlab's `fmincon` function using the interior point solver.

We did our calculations in Matlab 2015b on an Intel Core i7-5600U processor with 2.6 GHz and 8GB RAM.

Matlab's interior point solver took 16904 to 30690 seconds to solve the

lexicographic optimization problem for our four cases. Within a maximum of 1000 iterations per optimization level it yielded solutions which were slightly infeasible (maximal constraint violation of $\mathcal{O}(10^{-8})$ - $\mathcal{O}(10^{-6})$). Due to the nature of the evaluation functions this infeasibility had no major effect on the dose values. These solutions were still acceptable from a medicinal point of view.

For a detailed comparison of computation time see Table 6.

Table 6: Computation time in seconds used by Matlab’s interior point solver (IP), the unsuperiorized projection solver, the projection solver using robust parameters ($\hat{K}, \hat{\Lambda}$) (RobustSProj), and the projection solver using fast parameters ($\hat{K}, \hat{\Lambda}$) (FastSProj).

	IP	Proj	RobustSProj	FastSProj
case 1	30689.8	55.3	50.8	49.8
case 2	17649.8	14.4	12.8	9.5
case 3	26596.2	13.9	13.6	12.9
case 4	16903.9	14.8	13.0	13.5

All iterates produced by the projection solver are feasible by construction of the algorithm. This enables us to decide how much computational effort we want to put into the calculations - of course at the expense of mathematical solution quality.

For our calculations we chose the maximum number $n_{max} = 10^3$ of simultaneous projections, to solve a given CFP for the projection solver. When comparing the resulting solutions to the ones produced by Matlab’s solver (see Tables 7 and 8) it becomes apparent that they are suboptimal. Solutions with equal or smaller objective function values can be found by the projection solver, too, if n_{max} is raised.

Table 7: Objective funtion values (rounded to precision 10^0 if ≥ 1) at the solution x^* found by Matlab’s interior point solver

	$\phi_1(x^*)$	$\phi_2(x^*)$	$\phi_3(x^*)$	maximal constraint violation
case 1	$9.1529 \cdot 10^{-8}$	425	145	$9.1419 \cdot 10^{-6}$
case 2	$3.6419 \cdot 10^{-7}$	363	795	$7.2779 \cdot 10^{-6}$
case 3	0.3834	172	371	$9.8338 \cdot 10^{-7}$
case 4	0.4778	66	382	$9.6511 \cdot 10^{-8}$

Table 8: Objective funtion values (rounded to precision 10^0 if ≥ 1) at the solution x^* found by the superiorized projection solver with robust superiorization parameters $(\check{K}, \check{\Lambda})$ and $n_{max} = 10^3$.

	$\phi_1(x^*)$	$\phi_2(x^*)$	$\phi_3(x^*)$	maximal constraint violation
case 1	$\leq 1.1 \cdot 10^{-8}$	720	97	$\leq 10^{-8}$
case 2	$\leq 1.1 \cdot 10^{-8}$	502	1725	$\leq 10^{-8}$
case 3	$\leq 1.1 \cdot 10^{-8}$	1477	1918	$\leq 10^{-8}$
case 4	$\leq 1.1 \cdot 10^{-8}$	1318	851	$\leq 10^{-8}$

The DVH curves of case 1 in Figure 5 show that the additional 30639 seconds of computation time the Matlab solver takes compared to RobustSProj result in an improvement of 3 Gy in the average dose value of the LDP (whose dose evaluation function is contained in I_2) at the cost of 2 Gy increase in the average dose value of non-tumor tissue 1 (whose dose evaluation function is contained in I_3 and thus considered less important) .

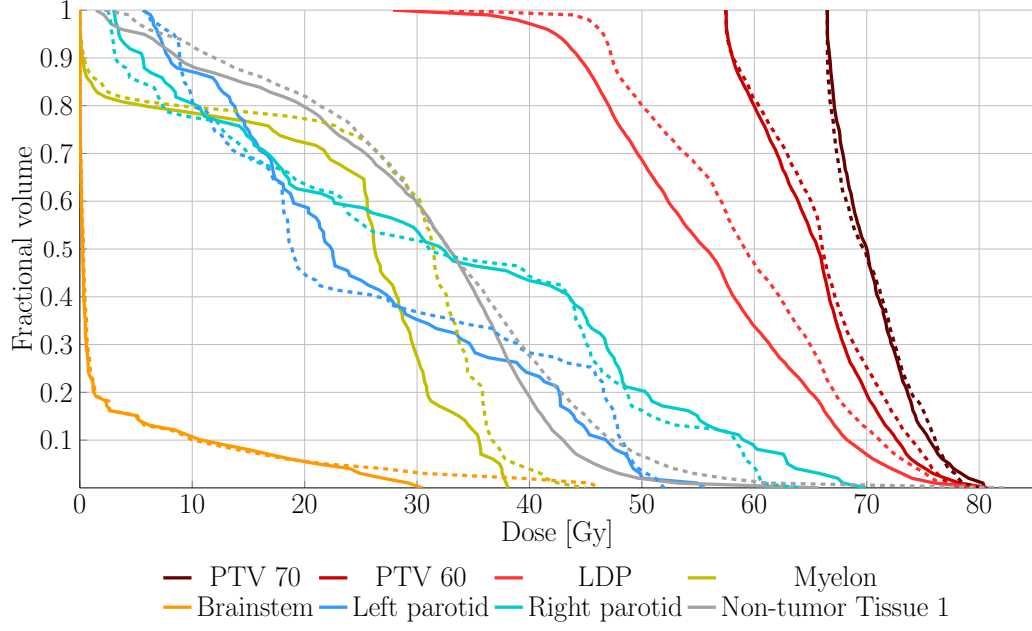


Figure 5: DVH comparison of solution doses of superiorized projection solver using superiorization parameters $(\tilde{K}, \tilde{\Lambda})$ and Matlab's interior point solver (dashed) for case 1.

For case 1, the Matlab solver yields a solution with has a lower evaluation function value corresponding to the LDP than the projection solver. Because in case 1 the LDP has a relatively big weight w_i in (2.3), this difference is amplified. We observed similar behavior with the other IMRT cases.

Concerning the question of performance compared to other solvers, we observe that the projection solver with $n_{max} = 10^3$ offers a fast way to reach a solution that is in many aspects already as good as the solution produced by Matlab's solver. If one is willing to invest additional hours of computation time, however, Matlab's solver offers solutions with lower objective function values.

5 Conclusion

We demonstrated the concept of superiorization in combination with an MCO method on a simple example as well as on four IMRT head neck cases. We showed that for LO we can speed up the optimization process by providing a

better starting point for the next level of optimization. We found a common set of superiorization parameters which yielded robust results for all four cases.

6 Acknowledgments

We thank the Department of Radiation Oncology (Head: Prof. Dr. C. Belka) of the Ludwig-Maximilians University (LMU), Munich, Germany, for providing the clinical data for our calculations and visualizations.

References

- [1] H. H. Bauschke and V. R. Koch, Projection Methods: Swiss Army Knives for Solving Feasibility and Best Approximation Problems with Halfspaces, *Contemporary Mathematics* **636** (2015), 1–40.
- [2] D. Butnariu, R. Davidi, G. T. Herman and I. G. Kazantsev, Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, *IEEE Journal of Selected Topics in Signal Processing* **1** (2007), 540–547.
- [3] Y. Censor, Weak and strong superiorization: Between feasibility-seeking and minimization, *Analele Stiintifice ale Universitatii Ovidius Constanta-Seria Matematica* **23** (2015), 41–54.
- [4] Y. Censor, W. Chen, P. L. Combettes, R. Davidi and G. T. Herman, On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints, *Computational Optimization and Applications* **51** (2012), 1065–1088.
- [5] Y. Censor, R. Davidi and G. T. Herman, Perturbation resilience and superiorization of iterative algorithms, *Inverse Problems* **26** (2010), 12pp.
- [6] Y. Censor, R. Davidi, G. T. Herman, R. W. Schulte and L. Tretuashvili, Projected subgradient minimization versus superiorization, *Journal of Optimization Theory and Applications* **160** (2014), 730–747.
- [7] G. Cimmino, Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari, *La Ricerca Scientifica* (Roma) **1** (1938), 326–333.

- [8] C. A. Coello Coello, A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques, *Knowledge and Information Systems* **1** (1999), 269–308.
- [9] R. Davidi, Y. Censor, R. W. Schulte, S. Geneser, and L. Xing, Feasibility-seeking and superiorization algorithms applied to inverse treatment planning in radiation therapy, *Contemporary Mathematics* **636** (2015), 83–92.
- [10] L. T. Dos Santos, A parallel subgradient method for the convex feasibility problem, *Journal of Computational and Applied Mathematics* **18** (1987), 307–320.
- [11] G. Eichfelder, Scalarizations for adaptively solving multi-objective optimization problems, *Computational Optimization and Applications* **44** (2009), 249–273.
- [12] A. Gibali, K.-H. Küfer, P. Süss, A generalized projection-based scheme for solving convex constrained optimization problems, Technical report 2015.
- [13] G. T. Herman, Superiorization for image analysis, in: *Combinatorial Image Analysis*, Lecture Notes in Computer Science **8466**, Springer, 2014, 1–7.
- [14] G. T. Herman, E. Garduño, R. Davidi and Y. Censor, Superiorization: An optimization heuristic for medical physics, *Medical Physics* **39** (2012), 5532–5546.
- [15] A. N. Iusem and L. Moledo, A finitely convergent method of simultaneous subgradient projections for the convex feasibility problem, *Computational and Applied Mathematics* **5** (1986), 169–184.
- [16] K.-W. Jee, D. L. McShan and B. A. Fraass, Lexicographic ordering: intuitive multicriteria optimization for IMRT, *Physics in Medicine and Biology* **52** (2007), 1845–1861.
- [17] W. Jin, Y. Censor and M. Jiang, A heuristic superiorization-like approach to bioluminescence, *International Federation for Medical and Biological Engineering (IFMBE) Proceedings* **39** (2013), 1026–1029.

- [18] S. N. Penfold, R. W. Schulte, Y. Censor and A. B. Rosenfeld, Total variation superiorization schemes in proton computed tomography image reconstruction, *Medical Physics* **37** (2010), 5887–5895.
- [19] I.-P. Stanimirović, Compendious lexicographic method for multi-objective optimization, *FACTA UNIVERSITATIS, Series Mathematics and Informatics* **27** (2012), 55–66.
- [20] K. Izui, T. Yamada and S. Nishiwaki, A gradient-based multiobjective optimization technique using an adaptive weighting method, *10th World Congress on Structural and Multidisciplinary Optimization* May 19–24, 2013, Orlando, Florida, USA.
- [21] Cancer Care Ontario, Dose Objectives for Head and Neck IMRT Treatment Planning Recommendation Report, 2014, <https://www.cancercare.on.ca/common/pages/UserFile.aspx?fileId=300041>.
- [22] J. O. Deasy, A. I. Blanco, V. H. Clark, CERR: A computational environment for radiotherapy research, *Medical Physics* **30** (2003), 979–985.
- [23] K.-H. Küfer, M. Monz, A. Scherrer, P. Süß, F. Alonso, A. Azizi Sultan, T. Bortfeld and C. Thieke, Multicriteria optimization in intensity modulated radiotherapy planning, *Springer Optimization and Its Applications, (Handbook of Optimization in Medicine)* **30** (2009), 123–168.
- [24] A. M. Geoffrion, Proper efficiency and the theory of vector maximization, *Journal of Mathematical Analysis and Applications* **22** (1968), 618–630.
- [25] T. Long, M. Matuszak, M. Feng, B. A. Fraass, R. K. Ten Haken and H. E. Romeijn, Sensitivity analysis for lexicographic ordering in radiation therapy treatment planning, *Medical Physics* **39** (2012), 3445–3455.